THEORY OF COMPUTING, Volume 7 (2011), pp. 75–99
www.theoryofcomputing.org

Testing Linear-Invariant Non-Linear Properties

Arnab Bhattacharyya* Victor Chen† Madhu Sudan‡ Ning Xie§

Received: September 18, 2009; published: March 29, 2011.

Abstract: We consider the task of testing properties of Boolean functions that are invariant under linear transformations of the Boolean cube. Previous work in property testing, including the linearity test and the test for Reed-Muller codes, has mostly focused on such tasks for linear properties. The one exception is a test due to Green for "triangle freeness": a function $f : \mathbb{F}_2^n \to \{0,1\}$ has this property if $f(x), f(y), f(x+y)$ do not all equal 1, for any pair $x, y \in \mathbb{F}_2^n$.

Here we extend this test to a more systematic study of testing for linear-invariant non-linear properties. We consider properties that are described by a single forbidden pattern (and its linear transformations), i. e., a property is given by k points $v_1, \dots, v_k \in \mathbb{F}_2^k$ and $f : \mathbb{F}_2^n \to \{0,1\}$ has the property that if for all linear maps $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$ it is the case that $f(L(v_1)), \dots, f(L(v_k))$ do not all equal 1. We show that this property is testable if the underlying matroid specified by $v_1, \dots, v_k$ is a graphic matroid. This extends Green's result to an infinite class of new properties. Part of our main results was obtained independently by Král', Serra, and Venna [*Journal of Combinatorial Theory Series A*, 116 (2009), pp. 971–978].

*Supported in part by a DOE Computational Science Graduate Fellowship and by NSF grants 0514771, 0732334, and 0728645.

†Supported in part by NSF Award CCR-0514915.

‡Supported in part by NSF Award CCR-0514915.

§Supported in part by an Akamai Presidential Fellowship and NSF grant 0514771.

ACM Classification: F.2.2

AMS Classification: 68W20, 68Q25

Key words and phrases: property testing, Fourier analysis

DOI: 10.4086/toc.2011.v007a006

Our techniques extend those of Green and in particular we establish a link between the notion of "1-complexity linear systems" of Green and Tao, and graphic matroids, to derive the results.

1 Introduction

Property testing considers the task of testing, "super-efficiently," if a function $f : D \to R$ mapping a finite domain D to a finite range R essentially possesses some desirable property. Letting $\{D \to R\}$ denote the set of all functions from D to R, a *property* is formally specified by a family $\mathcal{F} \subseteq \{D \to R\}$ of functions. A *tester* has oracle access to the function f and should accept with high probability (this probability is called the *completeness* of the tester) if $f \in \mathcal{F}$ and reject also with high probability (this probability is called the *soundness* of the tester) functions that are *far* from $\mathcal{F}$, while making very few queries to the oracle for f. Here, distance between functions $f, g : D \to R$, denoted $\delta(f,g)$, is simply the probability that $f(x) \neq g(x)$ when x is chosen uniformly at random from D and $\delta(f,\mathcal{F}) = \min_{g\in\mathcal{F}}\{\delta(f,g)\}$. We say f is δ-*far* from $\mathcal{F}$ if $\delta(f,\mathcal{F}) \geq \delta$ and δ-*close* otherwise. The central parameter associated with a tester is the number of oracle queries it makes to the function f being tested. In particular, a property is called *(locally) testable* if there is a tester with query complexity that is a constant depending only on the distance parameter δ. Property testing was initiated by the works of Blum, Luby and Rubinfeld [16] and Babai, Fortnow and Lund [9] and was formally defined by Rubinfeld and Sudan [30]. The systematic exploration of property testing was initiated by Goldreich, Goldwasser, and Ron [19] who expanded the scope of property testing to combinatorial and graph-theoretic properties (all previously considered properties were algebraic). In the subsequent years, a rich collection of properties have been shown to be testable [5, 4, 1, 17, 29, 3, 2, 25, 24] and many property tests have ended up playing a crucial role in constructions of probabilistically checkable proofs [8, 7, 11, 23, 31].

The rich collection of successes in property testing raises a natural question: Why are so many different properties turning out to be locally testable? Are there some broad "features" of properties that make them amenable to such tests? Our work is part of an attempt to answer such questions. Such questions are best understood by laying out broad (infinite) classes of properties (hopefully some of them are new) and showing them to be testable (or characterizing the testable properties within the class). In this paper we introduce a new such class of properties, and show that (1) they are locally testable, and (2) that they contain infinitely many new properties that were not previously known to be testable.

The properties, and our results The broad scope of properties we are interested in are properties that view their domain D as a vector space and are invariant under linear transformations of the domain. Specifically, we consider the domain $D = \mathbb{F}_2^n$, the vector space of n-dimensional Boolean vectors, and the range $R = \{0,1\}$. In this setting, a property $\mathcal{F}$ is said to be *linear-invariant* if for every $f \in \mathcal{F}$ and linear map $L : \mathbb{F}_2^n \to \mathbb{F}_2^n$ we have that $f \circ L \in \mathcal{F}$. Specific examples of linear-invariant properties that were previously studied (especially in the Boolean setting) include that of linearity, studied by Blum et al. [16] and Bellare et al. [10], and the property of being a "moderate-degree" polynomial (also known as a Reed-Muller codeword) studied by Alon et al. [2].[1] While the tests in the above-mentioned works potentially used all features of the property being tested, Kaufman and Sudan [26] showed that the testability can be attributed principally to the linear-invariance of the property. However their setting only considers *linear* properties, i. e., $\mathcal{F}$ itself is a vector space over $\{0,1\}$ and this feature plays a key role in their results: It lends an algebraic flavor to all the properties being tested and plays a central role in their analysis.

We thus ask the question: Is linear-invariance sufficient[2] to imply testability, even if the property being tested is non-linear? The one previous work in the literature that gave examples of non-linear linear-invariant properties is Green [21] where a test for the property of being "triangle-free" was described. A function $f : \mathbb{F}_2^n \to \{0,1\}$ is said to be *triangle-free* if for every $x, y \in \mathbb{F}_2^n$ it is the case that at least one of $f(x), f(y), f(x+y)$ does not equal 1. The property of being triangle-free is easily seen to be linear-invariant and yet not linear. Green [21] showed that the natural test for this property does indeed work correctly, though the analysis is quite different from that of typical algebraic tests and is more reminiscent of graph-property testing. In particular, Green develops an algebraic regularity lemma to analyze this test. (We note that the example above is not the principal objective of Green's work, which is directed mostly at abelian groups D and R. The above example with $D = \mathbb{F}_2^n$ and $R = \{0,1\}$ is used mainly as a motivating example.)

Motivated by the above example, we consider a broad class of properties that are linear-invariant and non-linear. A property in our class is given by k vectors $v_1, \ldots, v_k$ in the k-dimensional space $\mathbb{F}_2^k$. (Throughout this paper we think of k as a constant.) These k vectors uniformly specify a family $\mathcal{F} = \mathcal{F}_{n;v_1,\ldots,v_k}$ for every positive integer n, containing all functions that, for every linear map $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$ take on the value 0 on at least one of the points $L(v_1), \ldots, L(v_k)$. (In Section 6 we consider an even more generalized class of properties where the forbidden pattern of values for f is not 1^k but some other string and show a limited set of cases where we can test such properties.) To see that this extends the triangle-freeness property, note that triangle-freeness is just the special case with $k = 3$ and $v_1 = \langle 100 \rangle$, $v_2 = \langle 010 \rangle$, $v_3 = \langle 110 \rangle$. Under different linear transforms, these three points get mapped to all the different triples of the form $x, y, x+y$ and so $\mathcal{F}_{n;v_1,v_2,v_3}$ equals the class of triangle-free functions.

Before giving a name to our class of functions, we make a quick observation. Note that the property specified by $v_1, \ldots, v_k$ is equivalent to the property specified by $T(v_1), \ldots, T(v_k)$ where T is a non-singular linear map from $\mathbb{F}_2^k$ to $\mathbb{F}_2^k$. Thus the property is effectively specified by the dependencies among $v_1, \ldots, v_k$ which are in turn captured by the matroid[3] underlying $v_1, \ldots, v_k$. This leads us to our nomenclature:

Definition 1.1 ($\mathcal{M}$-freeness)**.** Given a (binary, linear) matroid $\mathcal{M}$ represented by vectors $v_1, \ldots, v_k \in \mathbb{F}_2^k$, the property of being $\mathcal{M}$-*free* is given by, for every positive integer n, the family

$$\mathcal{F}_{\mathcal{M}} = \{ f : \mathbb{F}_2^n \to \{0,1\} \mid \forall \text{ linear } L : \mathbb{F}_2^k \to \mathbb{F}_2^n, \langle f(L(v_1)), \ldots, f(L(v_k)) \rangle \neq 1^k \}.$$

[1] In the literature, the term "low-degree polynomial" is typically used for polynomials whose degree is smaller than the order of the field. In the paper [2] the degrees considered are larger than the order of the field, but are best thought of as large constants. The phrase "moderate-degree" above describes this setting of parameters.

[2] Strictly speaking, here we implicitly assume that the property satisfies any other necessary conditions for testability, such as being "locally characterized" [26].

[3] For the sake of completeness we include a definition of matroids in Section 2.1. However, a reader unfamiliar with this notion may just use the word "matroid" as a synonym for a finite collection of binary vectors, for the purposes of reading this paper.

The property of being $\mathcal{M}$-free has a natural k-local test associated with it: Pick a random linear map $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$ and test that $\langle f(L(v_1)), \ldots, f(L(v_k)) \rangle \neq 1^k$. Analyzing this test turns out to be non-trivial, and indeed we only manage to analyze this in special cases.

Recall that a matroid $\mathcal{M} = \{v_1, \ldots, v_k\}$, $v_i \in \mathbb{F}_2^k$, forms a *graphic matroid* if there exists a graph G on k edges with the edges being associated with the elements $v_1, \ldots, v_k$ such that a set $S \subseteq \{v_1, \ldots, v_k\}$ has a linear dependency if and only if the associated set of edges contains a cycle. In this paper, we require that the graph G be simple, that is, without any self-loops or parallel edges. Our main theorem shows that the property $\mathcal{F}$ associated with a graphic matroid $v_1, \ldots, v_k \in \mathbb{F}_2^k$ is testable.

Theorem 1.2. *For a graphic matroid $\mathcal{M}$, the property of being $\mathcal{M}$-free is locally testable. Specifically, let $\mathcal{M} = \{v_1, \ldots, v_k\}$ be a graphic matroid. Then, there exists a function $\tau : \mathbb{R}^+ \to \mathbb{R}^+$ and a k-query tester that accepts members of $\mathcal{M}$-free functions with probability one and rejects functions that are ε-far from being $\mathcal{M}$-free with probability at least $\tau(\varepsilon)$.*

Our bound on τ is quite weak. We let $W(t)$ denote a tower of twos with height $\lceil t \rceil$. Our proof only guarantees that $\tau(\varepsilon) \geq W(\mathrm{poly}(1/\varepsilon))^{-1}$, a rather fast vanishing function. We do not know if such a weak bound is required for any property we consider.

We describe the techniques used to prove this theorem shortly (which shed light on why our bound on τ is so weak) but first comment on the implications of the theorem. First, note that for a graphic matroid it is more natural to associate the property with the underlying graph. We thus use the phrase G-free to denote the property of being $\mathcal{M}$-free where $\mathcal{M}$ is the graphic matroid of G. This terminology recovers the notion of being triangle-free, as in [21], and extends to cover the case of being k-cycle free (also considered in [21]). But it includes every other graph too!

For every integer $k \geq 3$, let C_k denote the cycle graph with k vertices and let K_k denote the complete graph on k vertices. Syntactically, Theorem 1.2 seems to include infinitely many new properties (other than being k-cycle free). However, this may not be true semantically. For instance the property of being triangle-free is essentially the same as being G-free for every G whose biconnected components are triangles. Indeed, prior to our work, it was not even explicitly noted whether being C_k-free is essentially different from being triangle-free. (By "essentially," we ask if there exist triangle-free functions that are *far* from being C_k-free.) It actually requires careful analysis to conclude that the family of properties being tested includes (infinitely-many) new ones. Our second theorem addresses this point.

Theorem 1.3. *The class of G-free properties include infinitely many distinct ones. In particular:*

1. *For every odd k, if f is C_{k+2}-free, then it is also C_k-free. Conversely, there exist functions g that are C_k-free but far from being C_{k+2}-free.*
2. *If $k \leq \ell$ and f is K_k-free, then it is also K_ℓ-free. On the other hand, if $k \geq 3$ and $\ell \geq \binom{k}{2} + 2$ then there exists a function g that is K_ℓ-free but far from being K_k-free.*

Techniques Our proof of Theorem 1.2 is based on Green's analysis of the triangle-free case [21]. To analyze the triangle-free case, Green develops a "regularity" lemma for groups, which is analogous to Szemerédi's regularity lemma for graphs. In our setting, Green's regularity lemma shows how, given any function $f : \mathbb{F}_2^n \to \{0,1\}$, one can find a subgroup H of $\mathbb{F}_2^n$ such that the restriction of f to almost all cosets of H is "regular," where "regularity" is defined based on the "Fourier coefficients" of f. (These notions are made precise in Section 3.1.)

This lemma continues to play a central role in our work as well, but we need to work further on this. In particular, a priori it is not clear how to use this lemma to analyze $\mathcal{M}$-freeness for *arbitrary* matroids $\mathcal{M}$. To extract a large feasible class of matroids we use a notion from a work of Green and Tao [22] of the complexity of a linear system (or matroids, as we refer to them). The "least complex" matroids have complexity 1, and we show that the regularity lemma can be applied to all matroids of complexity 1 to show that they are testable (see Section 3).

The notion of a 1-complex matroid is somewhat intricate, and a priori it may not even be clear that this introduces new testable properties. We show (in Section 4) that these properties actually capture all graphic matroids which is already promising. Yet this is not a definite proof of novelty, and so in Section 5 we investigate properties of graphic matroids and give some techniques to show that they are "essentially" different. Our proofs show that if two (binary) matroids are not "homomorphically" equivalent (in a sense that we define) then there is an essential difference between the properties represented by them.

Significance of problems/results We now return to the motivation for studying $\mathcal{M}$-free properties. Our interest in these families is mathematical. We are interested in broad classes of properties that are testable; and invariance seems to be a central notion in explaining the testability of many interesting properties. Indeed this was what led Kaufman and Sudan [26] to examine this notion explicitly in the context of algebraic functions. They considered families that were linear-invariant and *linear*, and our work is motivated by the quest to see if the latter part is essential.

In contrast to other combinatorial settings, linear-invariance counts on a (quantitatively) very restricted collection of invariances. Indeed the set of linear transforms is only quasi-polynomially large in the domain (which may be contrasted with the exponentially large set of invariances that need to hold for graph-properties). So the ability to test properties based on this feature is mathematically interesting and leads to the question: what kind of techniques are useful in these settings. Our work manages to highlight some of those (in particular, Green's regularity lemma).

Related work Král', Serra and Venna in [28], independently of us, established a variant of Theorem 1.2 using completely different techniques. In our language, they demonstrated a direct reduction from testing freeness of a graphic matroid in a function to testing freeness of a subgraph in a graph, at which point they could apply a (colored) regularity lemma for graphs.

In the additive combinatorics community, subsequent to Green's work [21], there has been a lot of work developing better understanding of the regularity lemma and associated technology. In particular, Bergelson, Tao and Ziegler [12, 33] have resolved the inverse conjecture for the Gowers norms over finite fields. However, to the best of our knowledge, there does not yet exist in the literature a concrete formulation of a regularity lemma with respect to higher-order Gowers norms over $\mathbb{F}_2$, akin to Green's regularity lemma.

Subsequent Work Following our work, Shapira in [32] and Král', Serra and Vena in [27] independently extended the reduction in [28] to show that $\mathcal{M}$-freeness for an arbitrary matroid $\mathcal{M}$ is testable, resolving one of our main open questions. They achieved this by reducing the testing of $\mathcal{M}$-freeness in functions to testing whether a hypergraph is free from a fixed sub-hypergraph and then applying existing powerful hypergraph regularity lemmas. We remark that our proofs are very differently styled from those in [28], [27] and [32], and in particular, our view on invariance leads us to develop techniques which show that syntactically different properties are indeed distinct. We are also naturally led to the study of non-monotone properties (see Section 2.2) which were not investigated in these works.

More recent work has extended these aspects of our work. Bhattacharyya, Grigorescu and Shapira [15] have shown that non-monotone matroid-freeness properties are testable when the matroid is restricted to be of complexity 1. Interestingly, their proof uses arithmetic regularity lemmas, like Green [21] and us, instead of reductions to hypergraph properties as in [28, 27, 32]. The question of removing the restriction to matroids of complexity 1 for non-monotone properties remains open. Separately, Bhattacharyya, Grigorescu, Nordström and Xie [14] have extended the hierarchy of syntactically distinct properties to non-monotone matroid-freeness properties. Along the way, they demonstrate necessary and sufficient combinatorial conditions that determine when two matroid-freeness properties are identical and when they are far apart from each other, provided the matroids satisfy some structural conditions. Finally, Chen, Sudan and Xie [18] have shown that one can use self-correction, instead of the regularity lemma, to prove testability of some (non-monotone) matroid-freeness properties, thus dramatically improving the bounds on the query complexity of these properties in terms of dependence on ε.

For a more detailed discussion on the relation of this paper to subsequent works, we refer the reader to a report that we wrote recently [13].

Organization of this paper In the following section (Section 2) we provide some basic facts on matroids and define a slightly broader class of properties that we can consider (including some non-monotone properties). We also define the notion of 1-complexity matroids which forms a central tool in our analysis of the tests. In Section 3 we show that for any 1-complexity matroid $\mathcal{M}$, $\mathcal{M}$-freeness is testable. In Section 4 we show that graphic matroids are 1-complexity matroids. Theorem 1.2 thus follows from the results of Sections 3 and 4. In Section 5 we prove that there are infinitely many distinct properties among G-free properties. Section 6 and Section 7 are devoted to results on testing some non-monotone properties as well as some "collapse" results showing that many non-monotone properties collapse to some simple classes of functions. Finally Section 8 contains some concluding remarks.

2 Additional definitions, results, and overview of proofs

In this section, we first provide some additional definitions, then describe some further results that we present in the paper and finally give an outline of proofs.

2.1 Matroids

For more background on matroid theory, we refer the reader to [35].

Definition 2.1 (Matroids)**.** A matroid $\mathcal{M}$ is a finite set S (called *ground set*) and a collection $\mathcal{F}$ of subsets of S (called *independent sets*) such that the followings hold:

1. $\emptyset \in \mathcal{F}$.

2. If $X \in \mathcal{F}$ and $Y \subseteq X$, then $Y \in \mathcal{F}$.

3. If X and Y are both in $\mathcal{F}$ with $|X| = |Y| + 1$, then there exists an $x \in X \setminus Y$ such that $Y \cup x \in \mathcal{F}$.

A subset of the ground set $\mathcal{F}$ that is not independent is called *dependent*. A maximal independent set—that is, an independent set which becomes dependent on adding any element of $\mathcal{F}$—is called a *basis* for the matroid. The collection of bases of $\mathcal{M}$ is denoted by $\mathcal{B}(\mathcal{M})$.

A matroid $\mathcal{M}$ on a ground set $S = \{x_1, \ldots, x_k\}$ is said to be *linear* if there exists a field $\mathbb{F}$ and vectors $v_1, \ldots, v_k \in \mathbb{F}^k$ such that some subset $\{x_i \mid i \in T\}$ indexed by $T \subseteq \{1, \ldots, k\}$ is independent if and only if the corresponding vectors $\{v_i \mid i \in T\}$ are linearly independent. A linear matroid is *binary* if $\mathbb{F}$ can be chosen to be the field $\mathbb{F}_2$ of order 2.

Let G be a graph, let S be its set of edges $E(G)$. For all $X \subseteq E(G)$, let $X \in \mathcal{F}$ if and only if X does not contain a cycle of G. Then $\mathcal{F}$ is the collection of independent sets of a matroid on S, called the *cycle matroid* or *graphic matroid* of the graph G and denoted by $\mathcal{M}(G)$. Graphic matroids are binary; for example, if $\mathcal{M}$ is the graphic matroid of a graph G and I is the incidence matrix with rows indexed by edges of G and columns indexed by vertices of G, then the rows of I form a set of vectors that represent $\mathcal{M}$ over $\{0,1\}$.

If $\mathcal{M}$ is a finite matroid, then we define the *dual matroid* $\mathcal{M}^*$ by taking the same ground set and calling a set a basis in $\mathcal{M}^*$ if and only if its complement is a basis in $\mathcal{M}$. It can be easily verified that $\mathcal{M}^*$ is indeed a matroid and that the dual of $\mathcal{M}^*$ is $\mathcal{M}$. A matroid $\mathcal{M}$ is called a *cographic matroid* if its dual matroid $\mathcal{M}^*$ is graphic.

2.2 Extensions to non-monotone families

We now generalize Definition 1.1 to a wider collection of forbidden patterns.

Definition 2.2 ($\mathcal{M}$-freeness)**.** Given $\Sigma \in \mathbb{F}_2^k$ and a binary matroid $\mathcal{M}$ represented by vectors $v_1, \ldots, v_k \in \mathbb{F}_2^k$, the property of being $(\mathcal{M}, \Sigma)$-*free* is given by, for every positive n, the family

$$\mathcal{F}_{(\mathcal{M},\Sigma)} = \{ f : \mathbb{F}_2^n \to \{0,1\} \mid \forall \text{ linear } L : \mathbb{F}_2^k \to \mathbb{F}_2^n, \langle f(L(v_1)), \ldots, f(L(v_k)) \rangle \neq \Sigma \}.$$

If for some linear $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$, $\langle f(L(v_1)), \ldots, f(L(v_k)) \rangle = \Sigma$, then we say f *contains* $(\mathcal{M}, \Sigma)$ *at* L. Also, to be consistent with Definition 1.1, we suppress mention of Σ when $\Sigma = 1^k$.

Recall that a property $\mathcal{P} \subseteq \{D \to \{0,1\}\}$ is said to be *monotone* if $f \in \mathcal{P}$ and $g \prec f$ implies $g \in \mathcal{P}$, where $g \prec f$ means that $g(x) \leq f(x)$ for all $x \in D$.

Observation 2.3. For a binary matroid $\mathcal{M}$, $(\mathcal{M}, \Sigma)$-freeness is a monotone property if and only if $\Sigma = 1^k$.

In addition to our main results (Theorems 1.2 and 1.3) on monotone properties, we also obtain local testability results for a limited class of non-monotone properties.

Theorem 2.4. *Let C_k denote the cycle on k vertices and let Σ be an arbitrary element of $\mathbb{F}_2^k$. Then there exists a function $\tau : \mathbb{R}^+ \to \mathbb{R}^+$ and a k-query tester that accepts members of $\mathcal{F}_{(C_k,\Sigma)}$ with probability* 1 *and rejects f that are ε-far from $\mathcal{F}_{(C_k,\Sigma)}$ with probability at least $\tau(\varepsilon)$.*

However, in sharp contrast to Theorem 1.3, we show that, unless Σ equals 0^k or 1^k, the class of (C_k,Σ)-freeness properties is not at all very rich semantically.

Theorem 2.5. *The class of properties* $\{\mathcal{F}_{(C_k,\Sigma)} \mid k \geq 3, \Sigma \neq 0^k, \Sigma \neq 1^k\}$ *is only finitely large.*

The goal of Theorem 2.4 is not to introduce new testable properties but rather to illustrate possible techniques for analyzing local tests that may lead to more classes of testable non-monotone properties.

2.3 Overview of proofs

We now give an outline of the proofs of our main theorems (Theorems 1.2 and 1.3), and also the extensions (Theorems 2.4 and 2.5).

Our claim in Theorem 1.2, that graphic matroid freeness properties are locally testable, is based on analyzing the structure of dependencies among elements of a graphic matroid. To this end, we first recall the classification of linear forms due to Green and Tao [22]. We require a minor modification of their definition since, for us, the structure of the linear constraints is described by elements of a matroid.

Definition 2.6 (Complexity)**.** Given a binary matroid $\mathcal{M}$ represented by $v_1,\ldots,v_k \in \mathbb{F}_2^k$, we say that $\mathcal{M}$ has *complexity c at coordinate i* if we can partition $\{v_j\}_{j\in[k]\setminus\{i\}}$ into $c+1$ classes such that v_i is not in the span of any of the classes. We say that $\mathcal{M}$ has *complexity c* if c is the minimum such that $\mathcal{M}$ has complexity c at coordinate i for all $i \in [k]$.

To avoid triviality, we will consider in this paper only matroids whose complexities are greater than zero. This is because a matroid $\mathcal{M}$ has complexity 0 if and only if $\mathcal{M}$ consists of a set of linearly independent vectors. It follows that the all-zero function is the only function that is $\mathcal{M}$-free.

The above definition makes sense because the span of a set of elements is not dependent on the specific basis chosen to represent the matroid. As a motivating example, consider the graphic matroid of C_k studied by Green [21]. It can be represented by $v_1 = e_1, v_2 = e_2, \ldots, v_{k-1} = e_{k-1}$ and $v_k = e_1 + \cdots + e_{k-1}$. We see then that the graphic matroid of C_k has complexity 1 because for every $i < k$, the rest of the matroid elements can be partitioned into two sets $\{e_j\}_{j\neq i}$ and $\{\sum_{j\in[k]} e_j\}$ such that v_i is not contained in the span of either set, and for $i = k$, any nontrivial partition of the remaining elements ensures that v_k does not lie in the span of either partition. In Section 4, we extend this observation about C_k to all graphs.

Lemma 2.7. *For all graphs G, the graphic matroid of G has complexity* 1.

Green and Tao [22] showed that if a matroid $\mathcal{M}$ has complexity c and if A is a subset of $\mathbb{F}_2^n$, then the number of linear maps $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$ such that $L(v_i) \in A$ for all $i \in [k]$ is controlled by the $(c+1)^{\text{th}}$ Gowers uniformity norm of A. Previously, Green [21] proved an arithmetic regularity lemma, which essentially states that any set $A \subseteq \mathbb{F}_2^n$ can be partitioned into subsets of affine subspaces such that nearly every partition is nearly uniform with respect to linear tests. We show in Section 3 how to combine these two results to obtain the following:

Lemma 2.8. *Given any binary matroid $\mathcal{M}$ represented by $v_1,\ldots,v_k \in \mathbb{F}_2^k$, if $\mathcal{M}$ has complexity* 1*, then there exists a function $\tau : \mathbb{R}^+ \to \mathbb{R}^+$ and a k-query tester that accepts members of $\mathcal{F}_{\mathcal{M}}$ with probability* 1 *and rejects f that are ε-far from $\mathcal{F}_{\mathcal{M}}$ with probability at least $\tau(\varepsilon)$, where* $\tau(\varepsilon) \geq 1/W(\mathrm{poly}(1/\varepsilon))$.

Theorem 1.2 directly follows from combining Lemma 2.7 and Lemma 2.8. In fact, Lemma 2.8 implies testability of all matroids that have complexity 1, not only those that are graphic. In Section 4, we give examples of binary matroids that have complexity 1 and yet are provably not graphic.

Theorem 1.3 provides a proper hierarchy among the graphical properties. Moreover, the containments $\mathcal{P}_1 \subsetneq \mathcal{P}_2$ in this hierarchy are shown to be "statistically proper" in the sense that we demonstrate functions f that are ε-far from $\mathcal{P}_1$ but are in $\mathcal{P}_2$. The theorem implies the following hierarchy:

$$\cdots \subsetneq C_{k+2}\text{-free} \subsetneq C_k\text{-free} \subsetneq \cdots \subsetneq C_3\text{-free} = K_3\text{-free} \subsetneq \cdots \subsetneq K_k\text{-free} \subsetneq K_{\binom{k}{2}+2}\text{-free} \subsetneq \cdots$$

Thus, the class of properties $\mathcal{F}_G$ does indeed contain infinitely many more properties than the cycle freeness properties considered by Green [21].

Both the hierarchy among the cyclic freeness properties and among the clique freeness properties are derived in Section 5 using a general technique. In order to show a statistically proper containment $\mathcal{M}_1$-free $\subsetneq \mathcal{M}_2$-free, we construct a function f that, by its definition, contains $\mathcal{M}_1$ at a large number of linear maps and so is far from being $\mathcal{M}_1$-free. On the other hand, the construction ensures that if f is also not $\mathcal{M}_2$-free, then there is a *matroid homomorphism* from $\mathcal{M}_2$ to $\mathcal{M}_1$. We define a matroid homomorphism from a binary matroid $\mathcal{M}_2$ to a binary matroid $\mathcal{M}_1$ to be a map from the ground set of $\mathcal{M}_2$ to the ground set of $\mathcal{M}_1$ which maps cycles to cycles. The separation between $\mathcal{M}_2$-freeness and $\mathcal{M}_1$-freeness is then obtained by proving that there do not exist any matroid homomorphisms from $\mathcal{M}_2$ to $\mathcal{M}_1$. This proof framework suffices for both the claims in Theorem 1.3 and is reminiscent of proof techniques involving graph homomorphisms in the area of graph property testing (see [6] for a survey).

Theorem 2.4 is the result of a more involved application of the regularity lemma. To deal with non-monotone properties, we employ a different "rounding" scheme inspired by the testability of non-monotone graph properties in [1]. Unlike Szemerédi's regularity lemma, a "strong form" of the arithmetic regularity lemma is not known, so we restrict our attention to cyclic matroids and exploit the additive structure of the pattern. Theorem 2.5 is based on a characterization theorem in Section 7 that classifies (C_k,Σ)-freeness properties into 9 classes when $\Sigma \neq 0^k, 1^k$.

3 Freeness of complexity 1 matroids is testable

In this section we prove Lemma 2.8. Before doing so, we fix our notation and provide a quick background on Fourier analysis. We will be working with abelian groups with additive notation. If H is a subgroup of G, the cosets of H are indicated by $g+H$, with $g \in G$ chosen to represent that coset. Let $f_{g+H} : H \to \{0,1\}$ denote f restricted to the coset $g+H$, defined by sending h to $f(g+h)$; that is, for every $h \in H, g \in G$, $f_{g+H}(h) := f(g+h)$. For $\sigma \in \{0,1\}$, we define $\mu_\sigma(f_{g+H}) := \Pr_{h\in H}[f_{g+H}(h) = \sigma]$ to be the density of σ in f restricted to coset $g+H$.

3.1 Fourier analysis and Green's regularity lemma

Definition 3.1 (Fourier Transform)**.** If $f : \mathbb{F}_2^n \to \{0,1\}$, then we define its Fourier transform $\widehat{f} : \mathbb{F}_2^n \to \mathbb{R}$ to be $\widehat{f}(\alpha) = \mathbb{E}_{x\in\mathbb{F}_2^n} f(x)\chi_\alpha(x)$, where $\chi_\alpha(x) = (-1)^{\sum_{i\in[n]} \alpha_i x_i}$. $\widehat{f}(\alpha)$ is called the *Fourier coefficient* of f at α, and the $\{\chi_\alpha\}_\alpha$ are the characters of $\mathbb{F}_2^n$.

It is easy to see that for $\alpha,\beta \in \mathbb{F}_2^n$, $\mathbb{E}\,\chi_\alpha \cdot \chi_\beta$ is 1 if $\alpha = \beta$ and 0 otherwise. Since there are 2^n characters, the characters form an orthonormal basis for functions on $\mathbb{F}_2^n$, and we have the Fourier inversion formula

$$f(x) = \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)\chi_\alpha(x)$$

and Parseval's identity

$$\sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)^2 = \mathop{\mathbb{E}}_{x}\left[f(x)^2\right].$$

We extend Definition 3.1 to functions $f : H \to \{0,1\}$ where H is a subgroup of $\mathbb{F}_2^n$, by using the fact that H is isomorphic to $\mathbb{F}_2^m$ for some $m \le n$ and applying Definition 3.1 relative to this isomorphism.

Next we turn to Green's arithmetic regularity lemma, the crux of the analysis of our local testing algorithm. Green's regularity lemma over $\mathbb{F}_2^n$ is a structural theorem for Boolean functions. It asserts that for every Boolean function, there is some decomposition of the Hamming cube into cosets, such that the function restricted to most of these cosets are uniform and pseudorandom with respect to the linear functions. An alternate and equivalent way is that no matter where we cut the Boolean cube by a hyperplane, the densities of f on the two halves of the cube separated by the hyperplane do not differ greatly. Formally, we say that a function is uniform if all of its nonzero Fourier coefficients are small.

Definition 3.2 (Uniformity)**.** For every $0 < \varepsilon < 1$, we say that a function $f :\ \mathbb{F}_2^n \to \{0,1\}$ is *ε-uniform* if for every $\alpha \neq 0 \in \mathbb{F}_2^n$, $\left|\widehat{f}(\alpha)\right| \le \varepsilon$.

Recall that we let $W(t)$ denote a tower of twos with height $\lceil t \rceil$. To obtain a partition of the Hamming cube that satisfies the required uniformity requirement, the number of cosets in the partition may be rather large. More precisely,

Lemma 3.3 (Green's Regularity Lemma)**.** *Let $f :\ \mathbb{F}_2^n \to \{0,1\}$. For every $0 < \varepsilon < 1$, there exists a subspace H of $G = \mathbb{F}_2^n$ of co-dimension at most $W(\varepsilon^{-3})$, such that $\Pr_{g\in G}[f_{g+H}$ is ε-uniform$] \ge 1-\varepsilon$.*

3.2 Testability of complexity 1 matroid freeness

The proposition below was proved in [22]. Collectively, statements capturing the phenomenon that expectation over certain forms are controlled by varying degrees of the Gowers norm are termed *generalized von-Neumann type theorems* in the additive combinatorics literature. In particular, as we only require the degree 2 Gowers norm of a function, which is equivalent to the ℓ_4 norm of the function's Fourier transform. The version we state here requires the functions f_i to be over $\mathbb{F}_2^n$ and possibly distinct; however as explained by Gowers and Wolf [20], both conditions can be easily satisfied.

Proposition 3.4 (implicit in [22])**.** Suppose a binary matroid $\mathcal{M} = \{v_1 \ldots, v_k\}$ has complexity 1 and let $f_1,\ldots,f_k : \mathbb{F}_2^n \to \{0,1\}$. Then

$$\mathop{\mathbb{E}}_{L:\mathbb{F}_2^k \to \mathbb{F}_2^n}\left[\prod_{i=1}^k f_i(L(v_i))\right] \le \min_{i\in[k]}\left(\sum_{\alpha\in\mathbb{F}_2^n} \widehat{f_i}(\alpha)^4\right)^{1/4}.$$

It is an easy deduction from Proposition 3.4 to see that if f is uniform, then the number of linear maps L where f has a $\mathcal{M}$-pattern is close to $\mathbb{E}[f]^m N^d$, where $N = 2^n$. Combining this observation with the regularity lemma, we prove Lemma 2.8.

Proof of Lemma 2.8. Consider a test that picks a linear map L uniformly at random from all linear maps from $\mathbb{F}_2^k \to \mathbb{F}_2^n$ and rejects iff for all $i \in [k]$, $f(L(v_i)) = 1$. Clearly the test has completeness one.

Now we analyze the soundness of this test. Suppose f is ε-far from being $\mathcal{M}$-free. We want to show that the test rejects with probability at least $\tau(\varepsilon)$, such that $\tau(\varepsilon) > 0$ whenever $\tau > 0$. Let $a(\varepsilon)$ and $b(\varepsilon)$ be two functions of ε that satisfy the constraint $a(\varepsilon) + b(\varepsilon) < \varepsilon$, we shall specify these two functions at the end of the proof. We now apply Lemma 3.3 to f to obtain a subspace H of G of co-dimension at most $W(a(\varepsilon)^{-3})$. Consequently, f restricted to all but at most $a(\varepsilon)$ fraction of the cosets of H are $a(\varepsilon)$-uniform. We define a reduced function $f^R : \mathbb{F}_2^n \to \{0,1\}$ as follows.

For each $g \in G$, if f restricted to the coset $g + H$ is $a(\varepsilon)$-uniform, then define

$$f^R_{g+H}(x) = \begin{cases} 0 & \text{if } \mu_1(f_{g+H}) \le b(\varepsilon) \\ f_{g+H} & \text{otherwise.} \end{cases}$$

Else, define $f^R_{g+H} = 0$.

Note that at most $a(\varepsilon) + b(\varepsilon)$ fraction of modification has been made to f to obtain f^R. Since f is ε-far from being $\mathcal{M}$-free, f^R has a $\mathcal{M}$-pattern at some linear map L. More precisely, for every $i \in [k]$, $f^R(L(v_i)) = 1$. Now consider the cosets $L(v_i) + H$. By our construction of f^R, we know that f restricted to each of these cosets is $a(\varepsilon)$-uniform and at least $b(\varepsilon)$ dense. We will count the number of linear maps $\phi : \mathbb{F}_2^k \to H$ such that f has a $\mathcal{M}$ pattern at $L + \phi$. Notice that the probability the test rejects is at least

$$2^{-k \cdot W(a(\varepsilon)^{-3})} \Pr_{\phi:\mathbb{F}_2^k \to H}\left[\forall i, f_{L(v_i)+H}(\phi(v_i)) = 1\right].$$

To lower-bound this rejection probability, it suffices to show that the probability

$$\Pr_{\phi:\mathbb{F}_2^k \to H}\left[\forall i, f_{L(v_i)+H}(\phi(v_i)) = 1\right]$$

is bounded below by at least some constant depending on ε. To this end, we rewrite this probability as

$$\mathop{\mathbb{E}}_{\phi:\mathbb{F}_2^k \to H}\left[\prod_{i \in [k]} f_i(\phi(v_i))\right], \tag{3.1}$$

where $f_i = f_{L(v_i)+H}$. By replacing each function f_i by $\mathbb{E} f_i + (f_i - \mathbb{E} f_i)$, it is easy to see that the above expression can be expanded into a sum of 2^k terms, one of which is $\prod_{i \in [k]} \mathbb{E} f_i$, which is at least $b(\varepsilon)^k$. For the other $2^k - 1$ terms, by applying Proposition 3.4 and using Parseval's identity, each of these terms is bounded above by $a(\varepsilon)^{1/2}$. So Equation (3.1) is at least $b(\varepsilon)^k - (2^k - 1)a(\varepsilon)^{1/2}$. To finish the analysis, we need to specify $a(\varepsilon), b(\varepsilon)$ such that $b(\varepsilon)^k - (2^k - 1)a(\varepsilon)^{1/2} > 0$ and $a(\varepsilon) + b(\varepsilon) < \varepsilon$. Both are satisfied by setting $a(\varepsilon) = (\varepsilon/4)^{2k}$, $b(\varepsilon) = \varepsilon/2$. Thus, the rejection probability is at least

$$\tau(\varepsilon) \ge 2^{-k(W((4/\varepsilon)^{6k})+2)} \cdot \varepsilon^k,$$

completing the proof. □

4 Graphic matroids have complexity 1

Here we prove that graphic matroids have complexity 1. While the proof is simple, we believe it sheds insight into the notion of complexity and shows that even the class of 1-complexity matroids is quite rich.

Proof of Lemma 2.7. Recall that throughout we are assuming G to be a simple graph. Fix an arbitrary edge e in G with vertices v_1 and v_2 as its two ends. We partition the remaining edges of G into two sets S_1 and S_2 such that, if an edge is incident to v_1 then it is in S_1 and otherwise, it is in S_2. Because G is simple, a cycle in G containing e must include an edge (apart from edge e) which is incident to v_1 and another edge (other than e) which is not incident to v_1. Therefore e is not in the span of either S_1 or S_2. This shows that the complexity of the matroid of G is at most 1. But since we assume that the complexity of matroid of G is greater than 0, therefore the graphic matroid of G has complexity 1. □

As we have seen earlier, Lemma 2.8 holds for any matroid of complexity 1. Hence, it is a natural question to ask whether there exist non-graphic matroids which have complexity 1. In the following we show that such matroids do exist. It is an open question to come up with a natural characterization of matroids having complexity 1.

First we make the following claim which follows immediately from the definition of cographic matroids and the notion of complexity.

Claim 4.1. *A cographic matroid $\mathcal{M}^*(G)$ has complexity 1 if and only if, for every edge $e \in E(G)$, there is a partition of $E(G) \setminus \{e\}$ into two disjoint sets A and B such that both of the subgraphs $(V(G),A)$ and $(V(G),B)$ are connected.*

Proposition 4.2. There is a matroid with complexity 1 that is not graphic.

Proof. Consider the cographic matroid of K_5. Embed K_5 in the plane as a pentagon and all its diagonals. Fix an outer edge e and partition the remaining 9 edges into two sets. One is the 4 outer edges and the other is the remaining 5 diagonal edges. Clearly both outer-edge set and diagonal-edge set make the five vertices connected. Therefore by Claim 4.1, the cographic matroid of K_5 is of complexity 1. On the other hand, by a theorem of Tutte [34], a matroid cannot be graphic if it contains $\mathcal{M}^*(K_5)$ as a minor, which $\mathcal{M}^*(K_5)$ clearly does. So, $\mathcal{M}^*(K_5)$ is an example of a non-graphic matroid that has complexity 1. □

We remark that not all cographic matroids have complexity 1. For example, the cographic matroid of $K_{3,3}$ cannot have complexity 1 because if we remove an edge from $K_{3,3}$, there do not remain enough edges to form two edge-disjoint connected graphs on 6 vertices, violating Claim 4.1.

5 Infinitely many monotone properties

In this section we prove Theorem 1.3, that there are infinitely many matroids for which the property of being $\mathcal{M}$-free are pairwise very different.

To do so we consider a pair of target matroids $\mathcal{M}_1$ and $\mathcal{M}_2$. Based on just the first matroid $\mathcal{M}_1$, we create a canonical function $f = f_{\mathcal{M}_1} : \mathbb{F}_2^n \to \{0,1\}$. We show, using a simple analysis, that this canonical function is far from being $\mathcal{M}_1$-free. We then show that if this function has an instance of $\mathcal{M}_2$ inside, then there is a "homomorphism" (in a sense we define below) from $\mathcal{M}_2$ to $\mathcal{M}_1$. Finally we show two different ways in which one can rule out homomorphisms between pairs of graphic matroids; one based on the odd girth of the matroids, and the other based on the maximum degree of $\mathcal{M}_1$. Together these ideas lead to proofs of distinguishability of many different matroids.

Definition 5.1. Given a binary matroid $\mathcal{M}$ represented by vectors $v_1,\ldots,v_k \in \mathbb{F}_2^k$, and integer $n \geq k$, let the canonical function $f = f_{\mathcal{M}} : \mathbb{F}_2^n \to \{0,1\}$ be given by $f(x,y) = 1$ if $x \in \{v_1,\ldots,v_k\}$ and 0 otherwise; where $x \in \mathbb{F}_2^k$ and $y \in \mathbb{F}_2^{n-k}$.

Claim 5.2. *Let $\mathcal{M}$ be a binary matroid with $v_i \neq 0$ for all $i \in \{1,\ldots,k\}$. Then $f_{\mathcal{M}}$ is $\frac{1}{2^k}$–far from being $\mathcal{M}$-free.*

Proof. Note that if we consider the linear map $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$ that sends x to $\langle x, 0\rangle$, then f contains $\mathcal{M}$ at L. So f is not $\mathcal{M}$-free. However we wish to show that any function that is 2^{-k}-close to f contains $\mathcal{M}$ somewhere. Fix a function g such that $\delta(f,g) = \delta < 2^{-k}$. We will show that g contains $\mathcal{M}$ somewhere.

For $i \in [k]$ let

$$\delta_i = \Pr_{y \in \mathbb{F}_2^{n-k}}[f(v_i,y) \neq g(v_i,y)] .$$

Note that $\sum_{i=1}^k \delta_i \leq 2^k \cdot \delta < 1$. Now consider a random linear map $L_1 : \mathbb{F}_2^k \to \mathbb{F}_2^{n-k}$, and its extension $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$ given by $L(x) = \langle x, L_1(x)\rangle$. For every non-zero x and in particular for $x \in \{v_1,\ldots,v_k\}$, we have $L_1(x)$ is distributed uniformly over $\mathbb{F}_2^{n-k}$. Thus, for any fixed $i \in [k]$, we have $\Pr_{L_1}[g(L(v_i)) \neq 1] \leq \delta_i$. By the union bound, we get that $\Pr_{L_1}[\exists i \text{ such that } g(L(v_i)) \neq 1] \leq \sum_i \delta_i < 1$. In other words, there exists a linear map L_1 (and thus L) such that for every i, $g(L(v_i)) = 1$ and so g contains $\mathcal{M}$ at L. □

We now introduce our notion of a "homomorphism" between binary matroids. (We stress that the phrase homomorphism is conjured up here and we are not aware of either this notion, or the phrase being used in the literature. We apologize for confusion if this phrase is used to mean something else.)

Definition 5.3 (Homomorphism)**.** Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be binary matroids given by $v_1,\ldots,v_k \in \mathbb{F}_2^k$ and $w_1,\ldots,w_\ell \in \mathbb{F}_2^\ell$. We say that $\mathcal{M}_2$ has a homomorphism to $\mathcal{M}_1$ if there is a map $\phi : \{w_1,\ldots,w_\ell\} \to \{v_1,\ldots,v_k\}$ such that for every set $T \subseteq [\ell]$ such that $\sum_{i\in T} w_i = 0$, it is the case that $\sum_{i \in T} \phi(w_i) = 0$.

For graphic matroids, the matroid-homomorphism from G to H is a map from the edges of G to the edges of H that ensures that cycles are mapped to even degree subgraphs of H.

Lemma 5.4. *If the canonical function $f_{\mathcal{M}_1}$ contains an instance of $\mathcal{M}_2$ somewhere, then $\mathcal{M}_2$ has a homomorphism to $\mathcal{M}_1$.*

Proof. Let $f = f_{\mathcal{M}_1}$ contain $\mathcal{M}_2$ at L. So $L : \mathbb{F}_2^\ell \to \mathbb{F}_2^n$ is a linear map satisfying $f(L(w_i)) = 1$ for every $i \in [\ell]$. Now consider the projection map $\pi : \mathbb{F}_2^n \to \mathbb{F}_2^k$ which sends $\langle x,y\rangle$ to x (where $x \in \mathbb{F}_2^k$ and $y \in \mathbb{F}_2^{n-k}$).

We claim that the map ϕ which sends x to $\pi(L(x))$ gives a homomorphism from $\mathcal{M}_2$ to $\mathcal{M}_1$. On the one hand ϕ is linear and so if $\sum_{i\in T} w_i = 0$, then we have $\sum_{i\in T} \phi(w_i) = \phi(\sum_{i\in T} w_i) = \phi(0) = 0$. On the other hand, we also have that $\phi(w_i) \in \{v_1,\ldots,v_k\}$. This is true since $f(L(w_i)) = 1$, which implies, by the definition of the canonical function f, that $\pi(L(w_i)) \in \{v_1,\ldots,v_k\}$. Thus ϕ satisfies the requirements of a homomorphism from $\mathcal{M}_2$ to $\mathcal{M}_1$. □

The above lemma now motivates the search for matroids $\mathcal{M}_2$ that are not homomorphic to $\mathcal{M}_1$. Proving non-homomorphism in general may be hard, but we give a couple of settings where we can find simple proofs, each addresses a different case of Theorem 1.3.

For a matroid $\mathcal{M}$, let its *odd girth*, denoted $\mathrm{og}(\mathcal{M})$, be the size of the smallest dependent set of odd cardinality, i. e., the size of the smallest odd set $T \subseteq [\ell]$ such that $\sum_{i\in T} w_i = 0$.

Lemma 5.5. *If $\mathcal{M}_2$ has a homomorphism to $\mathcal{M}_1$, then $\mathrm{og}(\mathcal{M}_2) \geq \mathrm{og}(\mathcal{M}_1)$.*

Proof. Let ϕ be a homomorphism from $\mathcal{M}_2$ to $\mathcal{M}_1$ and let $T \subseteq [\ell]$ denote the smallest odd dependent set of $\mathcal{M}_2$. Now let $T' \subseteq [k]$ be the set

$$T' = \left\{ j \in [k] \;\middle|\; |\{i \in T \mid \phi(w_i) = v_j\}| \text{ is odd} \right\}.$$

On the one hand, we have T' has odd cardinality; and on the other, we have $0 = \sum_{i\in T} \phi(w_i) = \sum_{j\in T'} v_j$. So T' is an odd dependent set in $\mathcal{M}_1$. The lemma follows since $|T| \geq |T'|$. □

For graphic matroids constructed from the odd cycle graph C_k, we have that its odd girth is just k and so the above lemmas combine to give that C_k-freeness is distinguishable from C_{k+2}-freeness, and this suffices to prove Part (1) of Theorem 1.3.

However the odd girth criterion might suggest that G-freeness for any graph containing a triangle might be equivalent. We next rule this possibility out.

Lemma 5.6. *Let $\mathcal{M}_1$ be the graphic matroid of the complete graph K_a on a vertices, and let $\mathcal{M}_2$ be the graphic matroid of K_b. Then, if $b \geq \binom{a}{2} + 2$, there is no homomorphism from $\mathcal{M}_2$ to $\mathcal{M}_1$.*

Proof. Assume otherwise and let ϕ be such a homomorphism. Fix any vertex of K_b and let $e_1, \ldots, e_{b-1}$ denote the $b-1$ edges incident to this vertex. By the pigeonhole principle, (since $b - 1 > \binom{a}{2}$) there must exist a pair of incident edges e_i and e_j such that $\phi(e_i) = \phi(e_j)$. But now let f denote the edge which forms a triangle with e_i and e_j. Since in K_b we have $e_i + e_j + f = 0$ (viewing these elements as vectors over $\{0, 1\}$), it must be that $\phi(f) = \phi(e_i) + \phi(e_j) = 0$ which is not an element of the ground set of $\mathcal{M}_1$. This yields the desired contradiction. □

We are now ready to prove Theorem 1.3.

Proof of Theorem 1.3. First note that C_{k+2}-free functions are also C_k-free. Informally, suppose a function f has a k cycle at point $x_1, \ldots, x_k$, i. e., $f(x_i) = 1$ at these points and $\sum_i x_i = 0$. Then f has a $k+2$ cycle at the points $x_1, x_1, x_1, x_2, \ldots, x_k$. (This informal argument can obviously be converted to a formal one once we specify the graphic matroids corresponding to C_k and C_{k+2} formally.)

On the other hand, if we take $\mathcal{M}_1$ to be the graphic matroid corresponding to C_{k+2} and f to be the canonical function corresponding to $\mathcal{M}_1$, then by Claim 5.2 it is 2^{-k}-far from $\mathcal{M}_1$-free, and by Lemmas 5.4 and 5.5 it does not contain $\mathcal{M}_2$, the graphic matroid of C_k.

For the second part of the theorem, note that every property that is G-free is also H-free if G is a subgraph of H. Thus K_k-free is contained in K_ℓ free if $k \leq \ell$. The proper containment can now be shown as above, using Claim 5.2 and Lemmas 5.4 and 5.6. □

6 Testing non-monotone properties

In this section we prove Theorem 2.4. (Readers may find it useful to recall the background material in Section 3.1.) We show that for non-monotone properties, i. e., when $\Sigma \neq 0^k$ or 1^k, the property of $(\mathcal{M},\Sigma)$-free is testable when the underlying graph is a cycle. However, as opposed to Section 5, the number of non-monotone properties associated with cycles is finite. In fact we give a complete characterization of these non-monotone properties in Section 7.

Proof of Theorem 2.4. Suppose we have oracle access to a function $f : \mathbb{F}_2^n \to \{0,1\}$. Consider the following k-query test T, which selects a linear map $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$ uniformly at random from all such possible linear maps. T has oracle access to f and queries f at the points $L(v_1),\dots,L(v_k)$. T rejects iff all of these points are evaluated to 1. If f is $(\mathcal{M},\Sigma)$-free, T never rejects and has completeness 1.

Now we analyze the soundness of T. Suppose that f is ε-far from being $(\mathcal{M},\Sigma)$-free. We want to show that T rejects with probability at least $\tau(\varepsilon)$, such that $\tau(\varepsilon) > 0$ whenever $\varepsilon > 0$.

Let $1/2 < \eta < 1$ be any constant, and $a(\varepsilon)$ and $b(\varepsilon)$ be functions of ε that satisfy the constraints $a(\varepsilon)+b(\varepsilon) < \varepsilon$ and $1-\eta > b(\varepsilon)$. We shall specify these two functions at the end of the proof.

Now let G denote $\mathbb{F}_2^n$. We apply Lemma 3.3 to f to obtain a subspace H of G of co-dimension at most $W(a(\varepsilon)^{-3})$. We define a reduced function $f^R : \mathbb{F}_2^n \to \{0,1\}$ as follows. We assume that Σ has at least two occurrences of 1. (Otherwise it has at least two occurrences of 0, and in the construction of f^R, we flip the roles of 1 and 0 when f_{g+H} is not uniform. The rest of the proof will proceed analogously, and we leave its verification to the readers.)

For each $g \in G$, if f restricted to the coset $g+H$ is $a(\varepsilon)$-uniform, then define

$$f^R_{g+H} = \begin{cases} 0 & \text{if } \mu_1(f_{g+H}) < b(\varepsilon) \\ 1 & \text{if } \mu_1(f_{g+H}) > 1-b(\varepsilon) \\ f_{g+H} & \text{otherwise.} \end{cases}$$

Otherwise, define

$$f^R_{g+H} = \begin{cases} 1 & \text{if } \mu_1(f_{g+H}) \geq \eta \\ 0 & \text{otherwise.} \end{cases}$$

Note that at most $a(\varepsilon)+b(\varepsilon)$ fraction of modification has been made to f to obtain f^R, so f^R is ε-close to f. By assumption, f is ε-far from $(\mathcal{M},\Sigma)$-free, so f^R has a $(\mathcal{M},\Sigma)$ pattern at some linear map $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$, i. e., for each $i \in [k]$, $f^R(L(v_i)) = \sigma_i$, where $\Sigma = \langle \sigma_1,\dots,\sigma_k \rangle$, and $\sigma_i \in \{0,1\}$. Now consider the cosets $L(v_i)+H$. By our choice of rounding, f restricted to each $L(v_i)+H$ is dense in the symbol σ_i, i. e., $\mu_{\sigma_i}(f_{L(v_i)+H}) \geq b(\varepsilon)$, since $1-\eta \geq b(\varepsilon)$. We want to show that there are many $(\mathcal{M},\Sigma)$ patterns spanning across these cosets. In particular, we restrict our attention to the relative number of $(\mathcal{M},\Sigma)$-patterns at linear maps of the form $L+\phi$, where ϕ maps linearly from $\mathbb{F}_2^k$ to H. Notice that the probability the test T rejects is at least

$$2^{-(k-1)W(a(\varepsilon)^{-3})} \cdot \Pr_{\phi:\mathbb{F}_2^k \to H}[\forall i \in [k], f_{L(v_i)+H}(\phi(v_i)) = \sigma_i]\,.$$

It suffices to show that the probability

$$\Pr_{\phi:\mathbb{F}_2^k \to H}[\forall i \in [k], f_{L(v_i)+H}(\phi(v_i)) = \sigma_i] \tag{6.1}$$

is bounded below by some constant depending only on ε. To this end, we divide our analysis into two cases, based on whether there is some $j \in [k]$ such that $f_{L(v_j)+H}$ is $a(\varepsilon)$-uniform or not.

Case 1: There exists some $j \in [k]$ such that $f_{L(v_j)+H}$ is $a(\varepsilon)$-uniform.

For each $i \in [k]$, define $f_i : H \to \{0,1\}$ to be $f_i = f_{L(v_i)+H} + \sigma_i + 1$. Note that by definition, $\mathbb{E} f_i \geq b(\varepsilon)$. We begin by arithmetize Equation (6.1) as

$$\mathop{\mathbb{E}}_{\phi:\mathbb{F}_2^k \to H}\left[\prod_{i\in[k]} f_i(\phi(v_i))\right].$$

Since $\mathcal{M}$ is a cyclic matroid, by Fourier expansion,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\phi:\mathbb{F}_2^k \to H}\left[\prod_{i\in[k]} f_i(\phi(v_i))\right] \\
&= \mathop{\mathbb{E}}_{x_1,\dots,x_{k-1}\in H}\left[f_1(x_1)\cdots f_{k-1}(x_{k-1})f_k(x_1+\cdots+x_{k-1})\right] \\
&= \mathop{\mathbb{E}}_{x_1,\dots,x_{k-1}\in H}\left[\sum_{\alpha_1\in H}\widehat{f_1}(\alpha_1)\chi_{\alpha_1}(x_1)\cdots\sum_{\alpha_{k-1}\in H}\widehat{f_{k-1}}(\alpha_{k-1})\chi_{\alpha_{k-1}}(x_{k-1})\sum_{\alpha\in H}\widehat{f_k}(\alpha)\chi_\alpha(x_1+\cdots+x_{k-1})\right] \\
&= \sum_{\alpha_1,\dots,\alpha_{k-1},\alpha\in H}\widehat{f_1}(\alpha_1)\cdots\widehat{f_{k-1}}(\alpha_{k-1})\widehat{f_k}(\alpha)\left(\mathop{\mathbb{E}}_{x_1}[\chi_{\alpha_1+\alpha}(x_1)]\right)\cdots\left(\mathop{\mathbb{E}}_{x_{k-1}}[\chi_{\alpha_{k-1}+\alpha}(x_{k-1})]\right) \\
&= \sum_{\alpha_1,\dots,\alpha_{k-1},\alpha\in H}\widehat{f_1}(\alpha_1)\cdots\widehat{f_{k-1}}(\alpha_{k-1})\widehat{f_k}(\alpha)\prod_{i=1}^{k-1}\delta(\alpha,\alpha_i) \\
&= \sum_{\alpha\in H}\prod_{i\in[k]}\widehat{f_i}(\alpha),
\end{aligned}$$

where in the second last step we use the notation of δ-function, which is defined by $\delta(\alpha,\alpha_i)=1$ if $\alpha_i=\alpha$ and 0 otherwise.

Using the facts that $\mathbb{E} f_i \geq b(\varepsilon)$, f_j is $a(\varepsilon)$-uniform, there exist two distinct indices $i_1, i_2 \neq j \in [k]$ (since $k \geq 3$), Cauchy-Schwarz, and Parseval's identity, respectively, we have

$$\begin{aligned}
\sum_{\alpha\in H}\prod_{i\in[k]}\widehat{f_i}(\alpha) &\geq b(\varepsilon)^k - \sum_{\alpha\neq 0\in H}\prod_{i\in[k]}\left|\widehat{f_i}(\alpha)\right| \\
&\geq b(\varepsilon)^k - a(\varepsilon)\sum_{\alpha\neq 0\in H}\prod_{i\in[k]\setminus\{j\}}\left|\widehat{f_i}(\alpha)\right| \\
&\geq b(\varepsilon)^k - a(\varepsilon)\sum_{\alpha\neq 0\in H}\left|\widehat{f_{i_1}}(\alpha)\right|\left|\widehat{f_{i_2}}(\alpha)\right|, \\
&\geq b(\varepsilon)^k - a(\varepsilon)\left(\sum_{\alpha\neq 0\in H}\widehat{f_{i_1}}(\alpha)^2\right)^{1/2}\left(\sum_{\alpha\neq 0\in H}\widehat{f_{i_2}}(\alpha)^2\right)^{1/2} \\
&\geq b(\varepsilon)^k - a(\varepsilon).
\end{aligned}$$

To finish the analysis, we need to specify $a(\varepsilon), b(\varepsilon)$ such that the constraints $a(\varepsilon)+b(\varepsilon) < \varepsilon$ and $1-\eta > b(\varepsilon)$ are satisfied. Let

$$b(\varepsilon) = (1-\eta)\cdot\varepsilon \quad \text{and} \quad a(\varepsilon) = \frac{1}{2}(1-\eta)^k\varepsilon^k,$$

we have that the rejection probability is at least

$$\tau(\varepsilon) \geq 2^{-(k-1)W(a(\varepsilon)^{-3})}(1-\eta)^k\varepsilon^k/2.$$

Case 2: No $j \in [k]$ exists such that $f_{L(v_j)+H}$ is $a(\varepsilon)$-uniform.

Since $\mathcal{M}$ is a cyclic matroid, it is not hard to see that Equation (6.1) is equal to

$$\Pr_{\substack{x_1,\ldots,x_k\in H\\ \sum_i x_i=0}}\left[\forall i\in[k], f_{L(v_i)}(x_i)=\sigma_i\right]. \tag{6.2}$$

Since Σ contains at least two occurrences of the symbol 1, we may assume without loss of generality that $\sigma_{k-1}=\sigma_k=1$. Fix $x_1,\ldots,x_{k-2}\in H$ such that $f_{L(v_i)}(x_i)=\sigma_i$. Let $z=\sum_{i=1}^{k-2}x_i$. Since $\eta > 1/2$, by the union bound we have

$$\begin{aligned}\Pr_{x\in H}[f_{L(v_{k-1})}(x)=f_{L(v_k)}(x+z)=1] &= 1-\Pr_{x\in H}[f_{L(v_{k-1})}(x)=0 \text{ or } f_{L(v_k)}(x+z)=0]\\ &\geq 1-2(1-\eta)\\ &> 0.\end{aligned}$$

Since for each $i\in[k]$, $f_{L(v_i)+H}$ is not $a(\varepsilon)$-uniform, by our choice of rounding, $\Pr_{x\in H}[f_{L(v_i)}(x_i)=\sigma_i]$ is at least $1-\eta$. By picking $x_1,\ldots,x_{k-2}$ uniformly at random from H, it is not hard to see that the rejection probability of the test is at least

$$\tau(\varepsilon)\geq 2^{-(k-1)W(a(\varepsilon)^{-3})}(1-\eta)^{k-2}(2\eta-1),$$

where $a(\varepsilon)=\frac{1}{2}(1-\eta)^k\varepsilon^k$. □

7 Characterization of cycle-free functions

In this section we consider the property of being $(\mathcal{M},\Sigma)$-free, where $\mathcal{M}$ is the matroid of the k-cycle. Syntactically these appear to be infinitely many different properties. We show that there are only finitely many distinct properties here when Σ is not equal to 0^k or 1^k. (As noted in Section 5, when $\Sigma = 1^k$, we do get infinitely many distinct properties.)

Remarks on notation. For any $x\in\mathbb{F}_2^k$ and any $y\in\mathbb{F}_2^\ell$, we use xy to denote the concatenation of x and y, which is an element in $\mathbb{F}_2^{k+\ell}$. If $x\in\mathbb{F}_2^k$ is a binary string, then we write $\mathrm{ones}(x)$ for the number of indexes i such that $x_i=1$, and similarly write $\mathrm{zeros}(x)$ for the number of zeros in x. Let $f:\mathbb{F}_2^n\to\{0,1\}$ be a Boolean function. The *complement* of f, denoted $\bar{f}$, is also a Boolean function such that $\bar{f}(x)=1-f(x)$ for every $x\in\mathbb{F}_2^n$. Let $\mathbf{0}$ (resp. $\mathbf{1}$) stand for the all-zero (resp. all-one) function over the Boolean cube $\mathbb{F}_2^n$. Finally, we use the following (standard) terminology to describe the distinct families of Boolean functions which we are going to refer to in our characterization of (C_k,Σ)-free functions.

- **Const** is the set of constant functions (i. e., $\mathbf{Const} = \{\mathbf{0}, \mathbf{1}\}$).
- **Lin** denotes the set of all linear functions, including the constant functions.[4] **coLin** denotes the complementary family, i. e., all functions whose complements are in **Lin**.
- $\mathcal{F}_{\mathbf{lin}}$ stands for the family of characteristic functions of linear subspaces and the 0 function, i. e.,

$$\mathcal{F}_{\mathbf{lin}} = \{\mathbf{0}\} \cup \{ f : \mathbb{F}_2^n \to \{0,1\} \mid f^{-1}(1) \text{ is a linear subspace of } \mathbb{F}_2^n \} .$$

  $\mathbf{co}\mathcal{F}_{\mathbf{lin}}$ is the complementary family.

- $\mathcal{F}_{\mathbf{aff}}$ stands for the family of characteristic functions of affine subspaces and the 0 function, i. e.,

$$\mathcal{F}_{\mathbf{aff}} = \{\mathbf{0}\} \cup \{ f : \mathbb{F}_2^n \to \{0,1\} \mid f^{-1}(1) \text{ is an affine subspace of } \mathbb{F}_2^n \} .$$

  $\mathbf{co}\mathcal{F}_{\mathbf{aff}}$ is the complementary family.

It turns out that for every $k \geq 3$ and every $\Sigma \neq 0^k, 1^k$, a (C_k, Σ)-free family is one of the seven families $\mathbf{Const}, \mathbf{Lin}, \mathbf{coLin}, \mathcal{F}_{\mathbf{lin}}, \mathbf{co}\mathcal{F}_{\mathbf{lin}}, \mathcal{F}_{\mathbf{aff}}, \mathbf{co}\mathcal{F}_{\mathbf{aff}}$.

Theorem 7.1. *For every $k \geq 3$ and every $\Sigma \neq 0^k, 1^k$, a (C_k, Σ)-free family is one of*

$$\mathbf{Const}, \mathbf{Lin}, \mathbf{coLin}, \mathcal{F}_{\mathbf{lin}}, \mathbf{co}\mathcal{F}_{\mathbf{lin}}, \mathcal{F}_{\mathbf{aff}}, \mathbf{co}\mathcal{F}_{\mathbf{aff}}.$$

Specifically:

1. *If* $\mathrm{zeros}(\Sigma)$ *and* $\mathrm{ones}(\Sigma)$ *are both even, the* $\mathcal{F}_{C_k,\Sigma} = \mathbf{Const}$.
2. *If* $\mathrm{ones}(\Sigma) > 1$ *is odd and* $\mathrm{zeros}(\Sigma)$ *is even, then* $\mathcal{F}_{C_k,\Sigma} = \mathbf{Lin}$. *Complementarily, if* $\mathrm{zeros}(\Sigma) > 1$ *is odd and* $\mathrm{ones}(\Sigma)$ *is even, then* $\mathcal{F}_{C_k,\Sigma} = \mathbf{coLin}$.
3. *If* $\mathrm{ones}(\Sigma) = 1$ *and* $\mathrm{zeros}(\Sigma)$ *is even, then* $\mathcal{F}_{C_k,\Sigma} = \mathcal{F}_{\mathbf{lin}}$. *Complementarily, if* $\mathrm{zeros}(\Sigma) = 1$ *and* $\mathrm{ones}(\Sigma)$ *is even, then* $\mathcal{F}_{C_k,\Sigma} = \mathbf{co}\mathcal{F}_{\mathbf{lin}}$.
4. *If* $\mathrm{ones}(\Sigma), \mathrm{zeros}(\Sigma) > 1$ *and are both odd, then* $\mathcal{F}_{C_k,\Sigma} = \mathcal{F}_{\mathbf{aff}}$.
5. *If* $\mathrm{zeros}(\Sigma) = 1$ *and* $\mathrm{ones}(\Sigma) > 1$ *is odd, then* $\mathcal{F}_{C_k,\Sigma} = \mathcal{F}_{\mathbf{aff}}$. *Complementarily, if* $\mathrm{ones}(\Sigma) = 1$ *and* $\mathrm{zeros}(\Sigma) > 1$ *is odd, then* $\mathcal{F}_{C_k,\Sigma} = \mathbf{co}\mathcal{F}_{\mathbf{aff}}$.

We begin with some simple facts and observations.

Fact 7.2 ([29])**.** *Let S be an affine subspace. Then x, y and z are all in S implies $x+y+z$ is also in S. Conversely, if for any triple x, y and z in S implying $x+y+z$ in S, then S is an affine subspace.*

This fact immediately gives the following observations.

Observation 7.3. A function $f : \mathbb{F}_2^n \to \{0,1\}$ is $(C_4, 1110)$-free if and only if f is in $\mathcal{F}_{\mathbf{aff}}$.

[4] One may think of linear functions here as being the codewords of the Hadamard codes.

In a similar fashion, the following fact provides a characterization of $(C_3, 100)$-free (and $(C_3, 110)$-free) functions.

Fact 7.4. *A function $f : \mathbb{F}_2^n \to \{0,1\}$ is $(C_3, 100)$-free if and only if f is the disjunction (OR) of linear functions (or the all 1 function). Consequently, f is $(C_3, 110)$-free if and only if f is in $\mathcal{F}_{\mathbf{lin}}$.*

Proof. Let $S = \{x \in \mathbb{F}_2^n \mid f(x) = 0\}$. If S is empty, then f is the all 1 function. Otherwise let x and y be any two elements in S (not necessarily distinct). Then if f is $(C_3, 100)$-free, it must be the case that $x+y$ is also in S. Thus S is a linear subspace in $\mathbb{F}_2^n$. Suppose the dimension of S is k with $k \geq 1$. Then there are k linearly independent vectors $a_1, \ldots, a_k \in \mathbb{F}_2^n$ such that $z \in S$ iff $(\langle z, a_1 \rangle = 0) \wedge \cdots \wedge (\langle z, a_k \rangle = 0)$. Therefore, by De Morgan's law, $f(z) = 1$ iff $z \in \bar{S}$ iff $(\langle z, a_1 \rangle = 1) \vee \cdots \vee (\langle z, a_k \rangle = 1)$, which is equivalent to the claim. Similar argument proves the characterization of $(C_3, 110)$-free functions. □

Observation 7.5. If $\Sigma \neq 1^k$ for some $k > 2$, then $(C_{k+2}, \Sigma 00)$-free $\subseteq (C_k, \Sigma)$-free. Similarly, if $\Sigma \neq 0^k$ for some $k > 2$, then $(C_{k+2}, \Sigma 11)$-free $\subseteq (C_k, \Sigma)$-free.

Proof. By symmetry, we only need to prove the first part. Let $f \in (C_{k+2}, \Sigma 00)$-free. Suppose $f \notin (C_k, \Sigma)$-free, then there exists a violating tuple, say, $\langle x_1, x_2, \ldots, x_j \rangle$ such that $\sum_{i=1}^{j} x_i = 0$ and

$$\langle f(x_1), f(x_2), \ldots, f(x_j) \rangle = \Sigma .$$

Since Σ is not an all 1 vector, there exists some k such that $f(x_k) = 0$. But then $\langle x_1, x_2, \ldots, x_j, x_k, x_k \rangle$ would be a violation tuple of pattern $(C_{k+2}, \Sigma 00)$, contradicting our assumption that the function f is in $(C_{k+2}, \Sigma 00)$-free. □

Observation 7.6. $(C_4, 0011)$-free equals the set of constant functions.

Proof. Clearly a constant function has no 0011 pattern. For the reverse inclusion, suppose f is $(C_4, 0011)$-free but not a constant function. Then there exist x and y such that $f(x) = 0$ $f(y) = 1$. Then

$$\langle f(x), f(x), f(y), f(y) \rangle = 0011 . \qquad \square$$

Proof of Theorem 7.1.

1. This follows from Observation 7.5 and Observation 7.6.

2. We only need to prove the first half of the claim, since the second half will then follow by symmetry. It is easy to check that, if $\mathrm{ones}(\Sigma) > 1$ is odd and $\mathrm{zeros}(\Sigma)$ is even, then $\mathbf{Lin} \subseteq \mathcal{F}_{C_k,\Sigma}$. To prove the other containment, note that by Fact 7.4 and Observation 7.5, $\mathcal{F}_{C_k,\Sigma}$ is contained in the disjunction of linear functions and in particular, we may assume f is $(C_5, 00111)$-free. So

$$f(x) = (\langle a_1, x \rangle = 1) \vee (\langle a_2, x \rangle = 1) \vee \cdots ,$$

where $a_1, a_2, \ldots,$ are non-zero, distinct and linearly independent vectors. Since a_1 and a_2 are linearly independent, there exist x_1, x_2 such that $\langle a_1, x_1 \rangle = \langle a_2, x_2 \rangle = 1$ while $\langle a_1, x_2 \rangle = \langle a_2, x_1 \rangle = 0$. Then $\langle f(0), f(0), f(x_1), f(x_2), f(x_1 + x_2) \rangle = 00111$. Therefore f cannot be the disjunction of more than one linear function, making it linear. Finally note that

$$\mathbf{Lin} \subseteq (C_{|\Sigma|}, \Sigma)\text{-free} \subseteq (C_5, 00111)\text{-free} \subseteq \mathbf{Lin} .$$

3. This follows from Fact 7.4 and Observation 7.5.

4. Let i and j be odd integers. If $f(x)$ is a linear function, then it is $(C_{i+j}, 0^i 1^j)$-free since j is odd. If $f(x)$ is the complement of a linear function, then it is $(C_{i+j}, 0^i 1^j)$-free since i is odd. So if f is an affine function, then it is $(C_{i+j}, 0^i 1^j)$-free. Now consider $(C_6, 000111)$-free. If f is $(C_6, 000111)$-free then the set $f^{-1}(1)$ forms an affine subspace (since f is also $(C_4, 0111)$-free.). Similarly the set $f^{-1}(0)$ forms an affine subspace (since f is also $(C_4, 0001)$-free) and so f is an affine function.

5. This follows from Observation 7.3 and Observation 7.5. □

8 Conclusions and future work

We introduced an infinite family of properties of Boolean functions and showed them to be testable. Unfortunately, we were only able to analyze the tests when the matroid $\mathcal{M}$ was graphic and the pattern was monochromatic. This raises a plethora of new problems that we describe below.

The first natural quest is to generalize the problem to the solution to the case when the matroid is arbitrary over $\{0,1\}$, and further to the case when the matroid is over other fields. We note that this seems to pose significant technical hurdles and indeed even the simple property of being free of the matroid $\{e_1, e_2, e_3, e_1+e_2, e_2+e_3, e_3+e_1, e_1+e_2+e_3\}$ (where e_1, e_2, e_3 are linearly independent vectors) is problematic. Subsequent to this work, Shapira [32] and Král' et al [27] have successfully resolved this question.

Next, it would be nice to extend the results to the case where the pattern Σ is an arbitrary binary string, as opposed to being monotone. We did manage to extend this in the special case where $\mathcal{M}$ is a cyclic matroid, but in this case the extension is not very interesting. We do feel that our proof techniques already capture some non-trivial other cases, but are far from capturing all cases, even for graphic matroids.

Extending the patterns further, there is no real reason to view the range as a field element, so a major generalization would be to consider matroids over arbitrary fields, and letting the range be some arbitrary finite set R where the forbidden pattern $\Sigma \in R^k$. (We don't believe there should be any major technical barriers in this step, once we are able to handle arbitrary 0/1 patterns Σ.) Finally, all the above problems consider the case of a single forbidden pattern (and its linear transformations). A more general setting would be the forbidden pattern consisting of a collection of strings in $\mathbb{F}_2^k$, i. e., $\Xi = \{\Sigma_1, \dots, \Sigma_\ell\}$ for some $\ell > 1$.

These properties were specified by a matroid $\mathcal{M}$ on k elements and a pattern $\Xi \subseteq \mathbb{F}_2^k$. However to capture the full range of linear-invariant non-linear properties that allow one-sided error local tests, we should also allow the conjunction of a constant number of constraints. We believe this could lead to a characterization of all linear-invariant non-linear properties that allow one-sided error local tests.

In a different direction, we feel that it would also be nice to develop richer techniques to show the distinguishability of syntactically different properties. For instance, even for the graphic case we don't have a good understanding of when two different graphs represent essentially the same properties, and when they are very different.

Acknowledgments

We are grateful to Kevin Matulef for suggesting this research direction. We thank Tali Kaufman and Swastik Kopparty for helpful discussions. We thank Asaf Shapira for drawing our attention to his preprint [32]. Finally we would like to thank the anonymous referees for their useful comments and suggestions.

References

[1] NOGA ALON, ELDAR FISCHER, ILAN NEWMAN, AND ASAF SHAPIRA: A combinatorial characterization of the testable graph properties: It's all about regularity. *SIAM J. Comput.*, 39:143–167, 2009. [doi:10.1137/060667177] 76, 83

[2] NOGA ALON, TALI KAUFMAN, MICHAEL KRIVELEVICH, SIMON LITSYN, AND DANA RON: Testing low-degree polynomials over GF(2). In *Proc. Approx. Random. Comb. Opt. (RANDOM '03)*, volume 2764 of *Lecture Notes in Computer Science*, pp. 188–199. Springer, 2003. [doi:10.1007/978-3-540-45198-3_17] 76, 77

[3] NOGA ALON, MICHAEL KRIVELEVICH, ILAN NEWMAN, AND MARIO SZEGEDY: Regular languages are testable with a constant number of queries. *SIAM J. Comput.*, 30(6):1842–1862, 2000. [doi:10.1137/S0097539700366528] 76

[4] NOGA ALON AND ASAF SHAPIRA: A characterization of the (natural) graph properties testable with one-sided error. In *Proc. 46th FOCS*, pp. 429–438. IEEE Comp. Soc. Press, 2005. [doi:10.1109/SFCS.2005.5] 76

[5] NOGA ALON AND ASAF SHAPIRA: Every monotone graph property is testable. In *Proc. 37th STOC*, pp. 128–137. ACM Press, 2005. [doi:10.1145/1060590.1060611] 76

[6] NOGA ALON AND ASAF SHAPIRA: Homomorphisms in graph property testing - a survey. Technical Report 085, Electron. Colloq. on Comput. Complexity (ECCC), 2005. [ECCC:TR05-085] 83

[7] SANJEEV ARORA, CARSTEN LUND, RAJEEV MOTWANI, MADHU SUDAN, AND MARIO SZEGEDY: Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998. [doi:10.1145/278298.278306] 76

[8] SANJEEV ARORA AND SHMUEL SAFRA: Probabilistic checking of proofs: A new characterization of NP. *J. ACM*, 45(1):70–122, 1998. [doi:10.1145/273865.273901] 76

[9] LÁSZLÓ BABAI, LANCE FORTNOW, AND CARSTEN LUND: Non-deterministic exponential time has two-prover interactive protocols. *Comput. Complexity*, 1(1):3–40, 1991. [doi:10.1007/BF01200430] 76

[10] MIHIR BELLARE, DON COPPERSMITH, JOHAN HÅSTAD, MARCOS A. KIWI, AND MADHU SUDAN: Linearity testing over characteristic two. *IEEE Trans. Inform. Theory*, 42(6):1781–1795, 1996. [doi:10.1109/18.556674] 76

[11] MIHIR BELLARE, ODED GOLDREICH, AND MADHU SUDAN: Free bits, PCPs, and nonapproximability—towards tight results. *SIAM J. Comput.*, 27(3):804–915, 1998. [doi:10.1137/S0097539796302531] 76

[12] VITALY BERGELSON, TERENCE TAO, AND TAMAR ZIEGLER: An inverse theorem for the uniformity seminorms associated with the action of $\mathbb{F}^{\omega}$. *Geom. Funct. Anal.*, 19:1539–1596, 2010. [doi:10.1007/s00039-010-0051-1] 79

[13] ARNAB BHATTACHARYYA, VICTOR CHEN, MADHU SUDAN, AND NING XIE: Testing linear-invariant non-linear properties: A short report. Technical Report 10-116, Electron. Colloq. on Comput. Complexity (ECCC), July 2010. [ECCC:TR10-116] 80

[14] ARNAB BHATTACHARYYA, ELENA GRIGORESCU, JAKOB NORDSTRÖM, AND NING XIE: Separations of matroid freeness properties. Technical Report 10-136, Electron. Colloq. on Comput. Complexity (ECCC), August 2010. [ECCC:TR10-136] 80

[15] ARNAB BHATTACHARYYA, ELENA GRIGORESCU, AND ASAF SHAPIRA: A unified framework for testing linear-invariant properties. In *Proc. 51st FOCS*, pp. 478–487. IEEE Comp. Soc. Press, 2010. [doi:10.1109/FOCS.2010.53] 80

[16] MANUEL BLUM, MICHAEL LUBY, AND RONITT RUBINFELD: Self-testing/correcting with applications to numerical problems. *J. Comput. System Sci.*, 47(3):549–595, 1993. [doi:10.1016/0022-0000(93)90044-W] 76

[17] CHRISTIAN BORGS, JENNIFER T. CHAYES, LÁSZLÓ LOVÁSZ, VERA T. SÓS, BALÁZS SZEGEDY, AND KATALIN VESZTERGOMBI: Graph limits and parameter testing. In *Proc. 38th STOC*, pp. 261–270. ACM Press, 2006. [doi:10.1145/1132516.1132556] 76

[18] VICTOR CHEN, MADHU SUDAN, AND NING XIE: Property testing via set-theoretic operations. In *Proc. 2nd Symp. Innov. Comput. Sci. (ICS '11)*, pp. 211–222, 2011. 80

[19] ODED GOLDREICH, SHAFI GOLDWASSER, AND DANA RON: Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, 1998. [doi:10.1145/285055.285060] 76

[20] TIMOTHY GOWERS AND JULIA WOLF: The true complexity of a system of linear equations. Technical Report 0711.0185, arXiv e-print archive, 2007. [arXiv:0711.0185] 84

[21] BEN GREEN: A Szemerédi-type regularity lemma in abelian groups, with applications. *Geom. Funct. Anal.*, 15(2):340–376, 2005. [doi:10.1007/s00039-005-0509-8] 77, 78, 79, 80, 82, 83

[22] BEN GREEN AND TERENCE TAO: Linear equations in primes. *Ann. of Math.*, 171(3):1753–1850, 2010. [doi:10.4007/annals.2010.171.1753] 79, 82, 84

[23] JOHAN HÅSTAD: Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001. [doi:10.1145/502090.502098] 76

[24] CHARANJIT S. JUTLA, ANINDYA C. PATTHAK, ATRI RUDRA, AND DAVID ZUCKERMAN: Testing low-degree polynomials over prime fields. In *Proc. 45th FOCS*, pp. 423–432. IEEE Comp. Soc. Press, 2004. [doi:10.1109/FOCS.2004.64] 76

[25] TALI KAUFMAN AND DANA RON: Testing polynomials over general fields. In *Proc. 45th FOCS*, pp. 413–422. IEEE Comp. Soc. Press, 2004. [doi:10.1109/FOCS.2004.64] 76

[26] TALI KAUFMAN AND MADHU SUDAN: Algebraic property testing: The role of invariance. In *Proc. 40th STOC*, pp. 403–412. ACM Press, 2008. [doi:10.1145/1374376.1374434] 77, 79

[27] DANIEL KRÁL', ORIOL SERRA, AND LLUÍS VENA: A removal lemma for systems of linear equations over finite fields. Technical Report 0809.1846, arxiv e-print archive, 2008. [arXiv:0809.1846] 79, 80, 94

[28] DANIEL KRÁL', ORIOL SERRA, AND LLUÍS VENA: A combinatorial proof of the removal lemma for groups. *J. Combin. Theory Ser. A*, 116(4):971–978, 2009. [doi:10.1016/j.jcta.2008.12.003] 79, 80

[29] MICHAL PARNAS, DANA RON, AND ALEX SAMORODNITSKY: Testing basic Boolean formulae. *SIAM J. Discrete Math.*, 16(1):20–46, 2003. [doi:10.1137/S0895480101407444] 76, 92

[30] RONITT RUBINFELD AND MADHU SUDAN: Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996. [doi:10.1137/S0097539793255151] 76

[31] ALEX SAMORODNITSKY AND LUCA TREVISAN: Gowers uniformity, influence of variables, and PCPs. In *Proc. 38th STOC*, pp. 11–20, New York, NY, USA, 2006. ACM Press. [doi:10.1145/1132516.1132519] 76

[32] ASAF SHAPIRA: Green's conjecture and testing linear-invariant properties. In *Proc. 41st STOC*, pp. 159–166. ACM Press, 2009. [doi:10.1145/1536414.1536438] 79, 80, 94, 95

[33] TERENCE TAO AND TAMAR ZIEGLER: The inverse conjecture for the Gowers norm over finite fields via the correspondence principle. *Analysis & PDE*, 3(1):1–20, 2010. [doi:10.2140/apde.2010.3.1] 79

[34] WILLIAM T. TUTTE: Matroids and graphs. *Trans. Amer. Math. Soc.*, 90:527–552, 1959. [doi:10.2307/1993185] 86

[35] DOMINIC J.A. WELSH: *Matroid Theory*. Academic Press Inc., London, 1976. 80

AUTHORS

Arnab Bhattacharyya
CSAIL MIT, Cambridge, MA
abhatt@mit.edu
http://web.mit.edu/abhatt/www/

Victor Chen
Princeton University, Princeton, NJ
vychen@princeton.edu
http://people.csail.mit.edu/victor/

Madhu Sudan
Microsoft Research NE, Cambridge, MA
madhu@mit.edu
http://people.csail.mit.edu/madhu/

Ning Xie
CSAIL MIT, Cambridge, MA
ningxie@csail.mit.edu
http://people.csail.mit.edu/ningxie/

ABOUT THE AUTHORS

ARNAB BHATTACHARYYA is a Ph. D. student at MIT, advised by Ronitt Rubinfeld. His research interests are algorithms (sublinear-time and otherwise) and extremal and additive combinatorics. He enjoys pigeonholing (his colleagues') arguments, chopping misbehaving objects into more well-behaved ones, and studying ZFC for surprise quizzes. Lately, he has been really yearning for the days when Federer was the dominant tennis player in the world.

VICTOR CHEN did his undergraduate work at the University of Texas at Austin, and he obtained his Ph. D. under the supervision of Madhu Sudan at MIT. His research interests include property testing and combinatorics. In his spare time, he enjoys scavenging for used CDs and watching foreign movies.

MADHU SUDAN received his Ph. D. from the University of California at Berkeley in 1992. From 1992 to 1997 he was a research staff member at IBM's Thomas J. Watson Research Center. From 1997-2009 he was a faculty member at the Electrical Engineering and Computer Science Department at the Massachusetts Institute of Technology. He is currently a Principal Researcher at Microsoft Research New England. His research has focussed on Probabilistic Checking of Proofs, List-Decoding, Property Testing, and Semantic Communication.

NING XIE is a Ph. D. student at MIT. His advisor is Ronitt Rubinfeld. His main research interest is property testing.